Iniziato	Thursday, 13 January 2022, 15:13
Stato	Completato
Terminato	Thursday, 13 January 2022, 15:39
Tempo impiegato	26 min. 23 secondi
Punteggio	14,00/15,00
Valutazione	**28,00** su un massimo di 30,00 (**93**%)

Domanda **1**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is the main reason for the *standardization* of numeric attributes?

Scegli un'alternativa:

a. Change the distribution of the numeric attributes, in order to obtain gaussian distributions

b. Map all the numeric attributes to a new range such that the mean is zero and the variance is one. ✔

c. Map all the nominal attributes to the same range, in order to prevent the values with higher frequency from having prevailing influence

d. Remove non-standard values

Your answer is correct.

La risposta corretta è: Map all the numeric attributes to a new range such that the mean is zero and the variance is one.

Domanda **2**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following *is not* an objective of feature selection

Scegli un'alternativa:

- a. Reduce time and memory complexity of the mining algorithms
- b. Reduce the effect of noise
- c. Select the features with higher range, which have more influence on the computations ✔
- d. Avoid the *curse of dimensionality*

Risposta corretta.

La risposta corretta è: Select the features with higher range, which have more influence on the computations

Domanda **3**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following types of data allows the use of the euclidean distance?

Scegli un'alternativa:

- a. Ordered data
- b. Document representations
- c. Points in a vector space ✔
- d. Transactional data

Your answer is correct.

La risposta corretta è: Points in a vector space

Domanda **4**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Given the definitions below:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

which of the formulas below computes the *recall* of a binary classifier?

Scegli un'alternativa:

a. TN / (TN + FP)

b. TP / (TP + FP)

c. TP / (TP + FN) ✔ This is also called *sensitivity*, or *hit rate*, which is the number of detected true positives divided by the total number of positives

d. (TP + TN) / (TP + FP + TN + FN)

Risposta corretta.

La risposta corretta è: TP / (TP + FN)

Domanda **5**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What is the *Gini Index?*

Scegli un'alternativa:

a. An impurity measure of a dataset alternative to the *Information Gain* and to the *Misclassification Index* ✔

b. An accuracy measure of a dataset alternative to the *Information Gain* and to the *Misclassification Index*

c. An impurity measure of a dataset alternative to *overfitting* and *underfitting*

d. A measure of the *entropy* of a dataset

Your answer is correct.

La risposta corretta è: An impurity measure of a dataset alternative to the *Information Gain* and to the *Misclassification Index*

Domanda **6**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

A Decision Tree is...

Scegli un'alternativa:

- a. A tree-structured plan of tests on single attributes to forecast the cluster
- b. A tree-structured plan of tests on multiple attributes to forecast the target
- c. A tree-structured plan of tests on single attributes to forecast the target ✔
- d. A tree-structured plan of tests on single attributes to obtain the maximum purity of a node

Risposta corretta.

La risposta corretta è: A tree-structured plan of tests on single attributes to forecast the target

Domanda **7**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is the main purpose of *smoothing* in Bayesian classification?

Scegli un'alternativa:

- a. Dealing with missing values
- b. Classifying an object containing attribute values which are missing from some classes in the training set ✔
- c. Classifying an object containing attribute values which are missing from some classes in the test set
- d. Reduce the variability of the data

Risposta corretta.

La risposta corretta è: Classifying an object containing attribute values which are missing from some classes in the training set

Domanda **8**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What measure is maximised by the Expectation Masimisation algirithm for clustering?

Scegli un'alternativa:

- a. The likelihood of an attribute, given the class label
- b. The likelihood of an example
- c. The *likelihood* the distributions, defined by the parameters found, given the data available ✔
- d. The support of a class

Your answer is correct.

La risposta corretta è: The *likelihood* the distributions, defined by the parameters found, given the data available

Domanda **9**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the statements below is true? (Only one)

Scegli un'alternativa:

- a. K-means works well also with datasets having a very large number of attributes
- b. K-means finds the number of clusters which gives the minimum distortion
- c. Sometimes k-means stops to a configuration which does not give the minimum distortion for the chosen value of the number of clusters. ✔
- d. K-means always stops to a configuration which gives the minimum distortion for the chosen value of the number of clusters.

Your answer is correct.

La risposta corretta è: Sometimes k-means stops to a configuration which does not give the minimum distortion for the chosen value of the number of clusters.

Domanda **10**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

After fitting DBSCAN with the default parameter values the results are: 0 clusters, 100% of noise points. Which will be your next trial?

Scegli una o più alternative:

- a. Decrease the radius of the neighborhood
- b. Reduce the minimum number of objects in the neighborhood and the radius of the neighborhood
- c. Reduce the minimum number of objects in the neighborhood ✔
- d. Increase the radius of the neighborhood ✔

Risposta corretta.

Le risposte corrette sono: Reduce the minimum number of objects in the neighborhood, Increase the radius of the neighborhood

Domanda **11**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following statements regarding the discovery of association rules is true? (One or more)

Scegli una o più alternative:

- a. The confidence of a rule can be computed starting from the supports of itemsets ✔
- b. The support of an itemset is anti-monotonic with respect to the composition of the itemset ✔
- c. The support of a rule can be computed given the confidence of the rule
- d. The confidence of an itemset is anti-monotonic with respect to the composition of the itemset

Your answer is correct.

Le risposte corrette sono: The confidence of a rule can be computed starting from the supports of itemsets, The support of an itemset is anti-monotonic with respect to the composition of the itemset

Domanda **12**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

How does *pruning* work when generating frequent itemsets?

Scegli un'alternativa:

- a. If an itemset is frequent, then none of its subsets can be frequent, therefore the frequencies of the subsets are not evaluated
- b. If an itemset is not frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated ✔
- c. If an itemset is frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated
- d. If an itemset is not frequent, then none of its subsets can be frequent, therefore the frequencies of the subsets are not evaluated

Risposta corretta.

La risposta corretta è: If an itemset is not frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated

Domanda **13**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

In *feature selection*, what is the Principal Component Analysis?

- a. A mathematical technique used to find the principal attributes which determine the classification process ✖
- b. A heuristic technique used to find a subset of the attributes which produces the same classifier
- c. A mathematical technique used to transform non numeric attributes into numeric attributes
- d. A mathematical technique used to transform a set of numeric attributes into a smaller set of numeric attributes which capture most of the variability in data

Your answer is incorrect.

La risposta corretta è:
A mathematical technique used to transform a set of numeric attributes into a smaller set of numeric attributes which capture most of the variability in data

Domanda **14**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

How can we measure the quality of a trained regression model?

a. Counting the number of values correctly forecast
b. With a confusion matrix
c. With a formula elaborating the difference between the forecast values and the true ones ✔
d. With precision, recall and accuracy

Your answer is correct.

La risposta corretta è:
With a formula elaborating the difference between the forecast values and the true ones

Domanda **15**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is different from the others?

Scegli un'alternativa:

a. Dbscan ✔ This is not a classification method
b. SVM
c. Decision Tree
d. Neural Network

Risposta corretta.

La risposta corretta è: Dbscan

Vai a...

Machine Learning - Python Lab ►